

DEUTSCHES
PATENTAMT(21) Aktenzeichen: P 34 16 238.0
(22) Anmeldetag: 2. 5. 84
(43) Offenlegungstag: 20. 12. 84

4,707,858

DE 34 16 238 A 1

(30) Unionspriorität: (32) (33) (31)

02.05.83 US 490701

(71) Anmelder:

Motorola, Inc., Schaumburg, Ill., US

(74) Vertreter:

Grünecker, A., Dipl.-Ing.; Kinkeldey, H., Dipl.-Ing.
Dr.-Ing.; Stockmair, W., Dipl.-Ing. Dr.-Ing. Ae.E. Cal
Tech; Schumann, K., Dipl.-Phys. Dr.rer.nat.; Jakob,
P., Dipl.-Ing.; Bezold, G., Dipl.-Chem. Dr.rer.nat.;
Meister, W., Dipl.-Ing.; Hilgers, H., Dipl.-Ing.;
Meyer-Plath, H., Dipl.-Ing. Dr.-Ing., Pat.-Anw., 8000
München

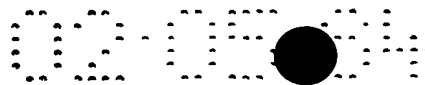
(72) Erfinder:

Fette, Bruce A., W. Del Campo Mesa, Ariz., US

(54) Extremschmalband-Übertragungssystem

Ein Übertragungssystem an dessen Enden jeweils eine Vorrichtung zum Analysieren menschlicher Sprache und zum Vergleichen jedes Wortes mit vorgespeicherten Wörtern zur Wort- und Sprechererkennung vorgesehen ist, wobei die Nachricht dann mit charakteristischen Eigenschaften der Stimme des Sprechers digitalisiert wird und ein Signal zur Übertragung mit einer Geschwindigkeit von etwa 75 Bit pro Sekunde gebildet wird, und eine Übertragung der digitalisierten Nachricht zu einem entfernten Terminal erfolgt, das diese Nachricht in eine gesprochene Nachricht in der synthetisierten Stimme des ursprünglichen Sprechers umwandelt.

DE 34 16 238 A 1



3416238

GRÜNECKER, KINKELDEY, STOCKMAIR & PARTNER

PATENTANWÄLTE
EUROPEAN PATENT ATTORNEYS

A. GRÜNECKER, DPL.-ING.
DR. H. KINKELDEY, DPL.-ING.
DR. W. STOCKMAIR, DPL.-ING./F. E. I. C. A. T. E. C. H.
DR. K. SCHUMANN, DPL.-PHYS.
P. H. JAKOB, DPL.-ING.
DR. G. BEZOLD, DPL.-CHEM.
W. MEISTER, DPL.-ING.
H. HILGERS, DPL.-ING.
DR. H. MEYER-PLATH, DPL.-ING.

8000 MÜNCHEN 22
MAXIMILIANSSTRASSE 55

30. April 1984
P 18 697

MOTOROLA, INC.
1303 E. Algonquin Road, Schaumburg,
Illinois 60196, USA

Extremschmalband-Übertragungssystem

P a t e n t a n s p r ü c h e

1. Extremschmalband-Übertragungssystem mit einem Wandler zum Umwandeln menschlicher Sprache in elektrische Signale, g e k e n n z e i c h n e t durch:

eine Analysiervorrichtung (15), die elektrische Signale vom Wandler (14) empfängt und eine Vielzahl von Signalen abgibt, die eine Vielzahl von Eigenschaften darstellen, die eine menschliche Stimme charakteri-

1 sieren;

eine Speichervorrichtung (20) in der Signale gespeichert sind, die eine Vielzahl gesprochener Wörter darstellen;

10 eine Worterkennungsvorrichtung (16), die mit der Analysiervorrichtung (15) und mit der Speichervorrichtung (20) zum Empfang von zumindest eines Teiles der Vielzahl von Signalen zum Vergleichen des empfangenen Teiles der Vielzahl von Signalen mit den gespeicherten Signalen verbunden ist, und Signale abgibt, die speziell gesprochene Wörter darstellen, und

15 eine Digitalumwandlungsvorrichtung, die mit der Worterkennungsvorrichtung (16) zum Empfang der spezifisch gesprochenen Wörter darstellenden Signale verbunden ist, zum Umwandeln der empfangenen Signale in eine Digitalform mit einer Geschwindigkeit von weniger
20 als 300 Bit pro Sekunde.

2. Extremschmalband-Übertragungssystem nach Anspruch 1, dadurch gekennzeichnet, daß die Analysiervorrichtung (15) eine Analysierschaltung (32) für
25 eine linear vorhersagbare Codierung aufweist.

3. Extremschmalband-Übertragungssystem nach Anspruch 1 oder 2, dadurch gekennzeichnet, daß die Worterkennungsvorrichtung (16) eine Vorrichtung (42, 43, 45) zum Erkennen des Beginns und des Endes eines
30 gesprochenen Wortes aufweist.

4. Extremschmalband-Übertragungssystem nach einem der Ansprüche 1 bis 3, dadurch gekennzeichnet, daß die Speichervorrichtung (20) Signale gespeichert
35 hat, die eine Vielzahl von Wörtern darstellen, die von einer Vielzahl unterschiedlicher Individuen gesprochen wurden, und daß eine Sprechererkennungsvorrichtung (18)

- 1 mit der Speichervorrichtung (20) und der Analysier-
vorrichtung (15) vorgesehen ist, die zumindest einen
Teil der Vielzahl von Signalen von der Analysiervor-
richtung (15) empfängt, die empfangenen Signale mit
5 den gespeicherten Signalen vergleicht und Signale ab-
gibt, die spezielle Wörter darstellen, die von einer
spezifischen, der unterschiedlichen Individuen ge-
sprochen wurden.
- 10 5. Extremschmalband-Übertragungssystem nach Anspruch 4,
dadurch g e k e n n z e i c h n e t, daß die Sprecher-
erkennungsvorrichtung (18) eine Schaltung zum Modifi-
zieren gespeicherter Wörter eines Individuum nach einer
Sprechererkennung aufweist.
- 15 6. Extremschmalband-Übertragungssystem nach Anspruch 5,
dadurch g e k e n n z e i c h n e t, daß die Analy-
siervorrichtung (15) eine Schaltung (32) zum Erhalten
von LPC-Koeffizienten einer linearen vorhersagbaren
20 Codierung und die Sprechererkennungsvorrichtung (18)
eine Schaltung zur Mittelwertbildung der LPC-Koeffizien-
ten aufweist.
- 25 7. Extremschmalband-Übertragungssystem nach Anspruch 6,
dadurch g e k e n n z e i c h n e t, daß die Sprecher-
erkennungsvorrichtung (18) eine Schaltung (20) zum Zu-
rückstellen einer Entscheidung bezüglich der Identi-
tät des Sprechers aufweist, wenn er der Vergleich ei-
nes gesprochenen Wortes mit gespeicherten Signalen,
30 die eine Vielzahl von einer Vielzahl von unterschiedli-
chen Individuen gesprochenen Wörtern darstellen, inner-
halb eines vorbestimmten Unsicherheitsbereichs liegt.
- 35 8. Extremschmalband-Übertragungssystem nach einem der
vorhergehenden Ansprüche, dadurch g e k e n n -
z e i c h n e t, daß die Signalumwandungsvorrichtung
(20) eine Einrichtung zum Umwandeln von Buchstaben
jedes spezifischen gesprochenen Wortes in ASCII Digi-

- 1 talcodierungen für eine Übertragung
9. Extremschmalband-Übertragungssystem nach einem der Ansprüche 4 bis 7, g e k e n n z e i c h n e t durch
5 eine Nachrichtenformatierungsvorrichtung (20), die mit der Sprechererkennungsvorrichtung (18) und der Worterkennungsvorrichtung (16) zum Formatieren jeder in den Wandler gesprochenen Nachricht in ein elektrisches Digitalsignal verbunden ist, das eine Vielzahl von
10 Bits enthält, die die Nachricht darstellen, sowie eine Vielzahl von Bits, die den Sprecher charakterisieren.
10. Extremschmalband-Übertragungssystem nach Anspruch 9, dadurch g e k e n n z e i c h n e t, daß die Nachricht-
15 formatierungsvorrichtung auch Bits verarbeitet, die Eigenschaften darstellen, die die Stimme des Sprechers charakterisieren.
11. Extremschmalband-Übertragungssystem nach Anspruch 10,
20 g e k e n n z e i c h n e t, durch eine Vorrichtung (20) zum Übertragen der Digitalsignale von der Umwandlungsvorrichtung zu einer entfernt angeordneten Einheit (12), eine Vorrichtung (20) zum Empfangen von Digitalsignalen und einer Synthesisiervorrichtung (22)
25 zum Umwandeln der Digitalsignale in die synthetisierte menschliche Sprache, die charakteristisch für die Stimme des Sprechers ist.
12. Extremschmalband-Übertragungsverfahren, g e k e n n -
30 z e i c h n e t durch die Schritte

Umwandeln von menschlicher Sprache in elektrische Signale,

35 Analysieren der elektrischen Signale, um eine Vielzahl von Signalen abzugeben, die eine Vielzahl von Eigenschaften darstellen, die eine menschliche Stimme charakterisieren,

- 1 Speichern von Signalen, die eine Vielzahl von ge-
sprochenen Wörtern darstellen,
- 5 Vergleichen zumindest einiger der Vielzahl von Sig-
nalen mit den gespeicherten Signalen, um spezifische
Wörter in der menschlichen Sprache zu bestimmen und
Signale abzugeben, die die spezifischen Wörter dar-
stellen, und
- 10 Umwandeln der abgegebenen Signale, die spezifische
Wörter darstellen in eine Digitalform mit einer Ge-
schwindigkeit geringer als 300 Bits pro Sekunde.
- 15 13. Verfahren nach Anspruch 12, g e k e n n z e i c h -
n e t durch Erkennen des Beginns und des Endes jedes
gesprochenen Worts vor dem Vergleichen.
- 20 14. Verfahren nach Anspruch 12 oder 13, dadurch g e -
k e n n z e i c h n e t, daß das Speichern eine Spei-
chern von Signalen umfaßt, die eine Vielzahl von von
einer Vielzahl unterschiedlicher Individuen gesproche-
nen Wörtern darstellen und daß das Vergleichen das Zu-
führen von Signalen umfaßt, die repräsentativ sind
25 für das individuelle Sprechen der spezifischen Wörter.
- 30 15. Verfahren nach einem der Ansprüche 12 bis 14, dadurch
g e k e n n z e i c h n e t, daß beim Analysieren
Koeffizienten mit linearer vorhersagbarer Codierung
erzeugt und die Koeffizienten vor dem Vergleichen ge-
mittelt werden.
- 35 16. Verfahren nach einem der Ansprüche 12 bis 15, dadurch
g e k e n n z e i c h n e t, daß das Vergleichen das
Zurückstellen einer Entscheidung bezüglich des indivi-
duellen Sprechens umfaßt, wenn der Vergleich eines
gesprochenen Wortes mit gespeicherten Signalen, die für
eine Vielzahl von durch eine Vielzahl von unterschied-

10

15

20

25

30

35

GRÜNECKER, KINKELDEY, STOCKMAIR & PARTNER

PATENTANWÄLTE
EUROPEAN PATENT ATTORNEYS

A. GRÜNECKER, DPL.-ING.
DR. H. KINKELDEY, DPL.-ING.
DR. W. STOCKMAIR, DPL.-ING., AER. INGENIEUR
DR. K. SCHUMANN, DPL.-PHYS.
P. H. JAKOB, DPL.-ING.
DR. G. BEZOLD, DPL.-CHEM.
W. MEISTER, DPL.-ING.
H. HILGERS, DPL.-ING.
DR. H. MEYER-PLATH, DPL.-ING.

8000 MÜNCHEN 22
MAXIMILIANSTRASSE 58

30. April 1984
P 18 697-57/to

MOTOROLA, INC.

1303 E. Algonquin Road, Schaumburg,

Illinois 60196, USA

Extremschmalband-Übertragungssystem

Beschreibung

In Übertragungssystemen ist es äußerst wünschenswert, Nachrichten mittels Sprache auszutauschen. Andererseits ist es erwünscht, digitale Schaltungen zu verwenden, da ein Großteil dieser Schaltungen auf einem einzigen integrierten Schaltungschip untergebracht werden können, was den erforderlichen Raum- und Energiebedarf wesentlich verringert. Digitale Darstellungen der menschlichen Sprache erfordern jedoch im allgemeinen eine verhältnismäßig große Bandbreite, so daß sie für viele Arten von Übertragungsmedien, etwa Telefonleitungen oder dergleichen, nicht geeignet sind. Die Bit-Übertragungsgeschwindigkeit (Bandbreite) von Nachrichten soll deshalb so niedrig wie möglich sein. Unter "Schmalband" wird üblicherweise eine Bit-Übertragungsgeschwindigkeit von etwa 2 000 Bits pro Sekunde verstanden. Bekannte Vorrichtungen arbeiten über

1 300 Bits pro Sekunde und alles, was darunter liegt, soll
als "Extremschmalband" bezeichnet werden.

Die vorliegende Erfindung betrifft ein Extremschmalband-
5 Übertragungssystem und ein Verfahren zum Nachrichtenaus-
tausch in einem extremen Schmalband, wobei menschliche
Sprache in elektrische Signale umgewandelt und analysiert
wird, so daß sich Signale ergeben, die Eigenschaften dar-
stellen, welche das spezielle menschliche Sprechen charak-
10 terisieren. Die Wörter der Nachricht werden dann mit Wör-
tern in einem Speicher verglichen, so daß das spezielle
~~Wort erkannt wird~~ und falls erwünscht auch der spezielle Spre-
cher, der dieses Wort ausgesprochen hat. Ein das speziel-
le Wort darstellendes Digitalsignal, das eine ASCII- oder
15 numerische Kodierung sein kann und die Position des Wor-
tes im Speicher angibt, wird mit Digitalsignalen kombi-
niert, die die Stimme des Sprechers charakterisieren, da-
mit sich eine Nachricht ergibt mit einer Bit-Geschwindig-
keit wesentlich unter 300 Bit pro Sekunde, wobei die Nach-
20 richt zu einem entfernten Endgerät übertragen wird. Die-
ses Endgerät synthetisiert die menschliche Stimme, so daß
die Nachricht derart ertönt, als wenn die ursprüngliche
Stimme sprechen würde. Verschiedene Verfahren und Einrich-
tungen dienen dazu, die korrekte Erkennung jedes Wortes
25 und des speziellen Sprechers zu gewährleisten einschließ-
lich einer Mittelwertbildung von LPC-Koeffizienten, Hin-
ausschieben einer Entscheidung bezüglich der Identität des
Sprechers, wenn der Vergleich der gesprochenen mit den ge-
speicherten Wörtern innerhalb eines vorbestimmten Unsicher-
30 heitsbereichs liegt und Modifizieren beziehungsweise auf
den neuesten Stand Bringen der gespeicherten Wörter eines
individuellen Sprechers, nachdem dieser erkannt wurde.

Der Erfindung liegt die Aufgabe zugrunde, ein neues und
35 verbessertes Extremschmalband-Übertragungssystem anzuge-
ben.

- 1 Ferner soll ein verbessertes Verfahren des Nachrichtenaustausches mittels Extremschmalband aufgezeigt werden.

An der empfangenden Endstation soll eine Stimme synthetisiert werden, die gleich derjenigen des ursprünglichen Sprechers ist.

Die Erkennung des Sprechers soll äußerst genau erfolgen.

- 10 Ein Ausführungsbeispiel der Erfindung wird nachstehend unter Bezugnahme auf die Zeichnung beschrieben. Es zeigen

- 15 Figur 1 ein vereinfachtes Blockschaltbild eines Extremschmalbandnachrichten- oder Übertragungssystems der Ausführungsform der Erfindung,
- Figur 2 ein Blockschaltbild der LPC-Analysiereinheit des Systems nach Figur 1,
- 20 Figur 3 ein Blockschaltbild der CPU-Einheit des Systems nach Figur 1,
- Figur 4 ein Blockschaltbild der Worterkennungseinheit des Systems nach Figur 1,
- Figur 5 ein Blockschaltbild der Synthetisiereinheit des Systems nach Figur 1,
- 25 Figur 6 ein Flußdiagramm zur Veranschaulichung des Beginns und der Beendigung einer Wortidentifikation in der Worterkennungseinheit der Figur 4,
- Figur 7 ein Flußdiagramm beziehungsweise ein Syntaxbaum bestimmt für militärische Zwecke und
- 30 Figur 8 vier typische Anzeigebilder im Zusammenhang mit dem Flußdiagramm der Figur 7.

Figur 1 zeigt das Extremschmalband-Übertragungssystem gemäß dem Ausführungsbeispiel der Erfindung. Ein Ortsterminal 10 und ein entferntes Terminal 12 sind über ein geeignetes Mittel, etwa Telefonleitungen oder dergleichen, ver-

1 bunden. Das Ortsterminal 10 weist ein Mikrofon 14 zum Um-
wandeln der menschlichen Sprache in elektrische Signale
in üblicher Art auf und ist mit einer LPC-Analysierein-
heit 15 und einer Worterkennungseinheit 16 verbunden.

5 LPC-Analyse bedeutet Analyse einer linearen vorhersagba-
ren Kodierung. Die LPC-Analysiereinheit oder -schaltungs-
platte 15 ist an eine zentrale Verarbeitungseinheit CPU 18
angeschlossen, die wiederum mit einem Rechner 20 in Ver-
bindung steht, der ein Tastenfeld, einen Austauschplatten-
10 speicher (Floppydiscspeicher) und eine Sichtanzeige auf-
weist. Die Worterkennungseinheit 16 ist mit dem Perso-
nalrechner 20 und eine Synthetisiereinheit oder -schal-
tungsplatte 22 ist ebenfalls mit dem Rechner 20 verbunden.
Der Ausgang der Synthetisiereinheit 22 liegt an Kopfhö-
15 rern 23 oder einem anderen Wandler geeigneter Art zum Um-
wandeln elektrischer Signale von der Synthetisierein-
heit 22 in Schall.

Figur 2 zeigt in größerer Einzelheit ein Blockschaltbild
20 der LPC-Analysiereinheit 15 in Form eines vollständigen
digitalen Sprachverarbeitungssystems, wie es im einzelnen
in der noch schwebenden US-Patentanmeldung mit der Be-
zeichnung "Digital Voice Processing System" und dem Akten-
zeichen 309 640 vom 8. Oktober 1981 beschrieben ist. Die
25 LPC-Analysiereinheit ist nur ein Teil des in Figur 2 ver-
anschaulichten Systems und ist im einzelnen in der
US-PS 4 378 469 erläutert. Das vollständige Verarbei-
tungssystem ist deshalb beschrieben, weil es einen Teil
der LPC-Analysiereinheit 15 darstellt und der Syntheti-
30 sierteil der Einheit 15 zur Synthetisierung der mensch-
lichen Stimme verwendet werden kann, so daß sie am ent-
fernten Terminal 12 wie das Sprechen eines Sprechers er-
tönt. Im vorliegenden System wird der Synthetisierer der
Einheit 15 nicht verwendet. Der Fachmann erkennt jedoch,
35 daß diese Einheit ohne weiteres an Stelle der Syntheti-
siereinheit 22 eingesetzt werden kann.

1 Gemäß Figur 2 werden Tonfrequenzsignale von dem Mikro-
fon 14 über eine AVR-Schaltung 25 mit automatischer Ver-
stärkungsregelung und ein Tiefpaßfilter 26 einer Abtast-
und Halteschaltung 28 zugeführt. Diese arbeitet mit einem
5 Analog-/Digitalwandler 30 zusammen, um für jede durch die
Abtast- und Halteschaltung 28 durchgeführte Abtastung eine
12 Bit-Digitaldarstellung abzugeben. Diese Digitalwerte
von dem A/D-Wandler 30 werden einer LPC-Analysiereinheit 32
zugeführt, die in der vorgenannten Patentschrift im ein-
10 zeln beschriebenen ist. Die LPC-Analysiereinheit 32 gibt
mehrere Signale ab, die unterschiedliche Eigenschaften
darstellen, die eine menschliche Stimme charakterisieren,
wie den Tonhöhenfrequenzbereich und eine Abschätzung der
vokalen Spurlänge sowie wahlweise einsetzbare zusätzliche
15 Eigenschaften, wie die glottale Erregungsform im Frequenz-
bereich und der Heiserkeitsgrad etc. Die Signale von der
LPC-Analysiereinheit 32 umfassen auch einen RMS-Durch-
schnittswert und eine vorbestimmte Anzahl von LPC-Koeffi-
zienten, nämlich in diesem Ausführungsbeispiel zehn. Alle
20 diese Signale von der LPC-Analysiereinheit 32 werden über
eine Schnittstelle 34 der CPU 18 zur Speicherung und Ver-
arbeitung zugeführt. Ein detaillierteres Blockschaltbild
der CPU 18 ist in Figur 3 gezeigt. Bei diesem Ausführungs-
beispiel ist die CPU 18 die im Handel erhältliche
25 CMT-68K-CPU. Da die in Figur 3 veranschaulichte CPU 18
im Handel erhältlich ist, kennt der Fachmann die Arbeits-
weise. Da alle Blöcke ausreichend definiert sind, soll de-
ren Funktion nicht im einzelnen beschrieben werden.

30 Obwohl die verschiedensten Einrichtungen als Worterken-
nungseinheit 16 verwendet werden können, kommt bei der
vorliegenden Ausführungsform die im Handel erhältliche
Einheit VRM102 zum Einsatz, die anhand der Figur 4 erläu-
tert wird. Die Tonfrequenzsignale vom Mikrofon 14 werden
35 an den Audioeingang angelegt und über einen Vorverstär-
ker 35 zum 16 Filter-Analysierer 37 geleitet. Der 16 Fil-
ter-Analysierer 37 führt grundsätzlich die Analysierfunkt-

tion der LPC-Analysiereinheit durch und der Fachmann erkennt, daß eine Worterkennungseinheit auch auf Signale der LPC-Analysiereinheit 15 basieren kann. Das Ausgangssignal des 16 Filter-Analysierers 37 wird über einen Gleichrichter 39 an einen 8 Bit-Analog-/Digitalwandler 40 angelegt. Dieser A/D-Wandler 40 ist mit einem 6802 Mikroprozessor 42, einem 4K-RAM-Speicher 43 und einem 4K-ROM-Speicher 45 verbunden. Die Worterkennungseinheit 16 besitzt auch mehrere Anschlüsse und Puffer zum Nachrichtenaustausch mit dem Personalrechner 20, dessen Funktion bekannt ist und hier nicht im einzelnen beschrieben wird.

Spektralamplituden des Gleichrichters 39 werden alle 5 ms durch den A/D-Wandler 40 ausgelesen. Das System mißt die Spektraldifferenz zwischen dem augenblicklichen Spektrum und dem Hintergrundrauschen. Überschreitet diese Differenz einen ersten Schwellenwert, dann markiert das System den möglichen Beginn eines Wortes und spektrale Abtastungen werden in dem "UNBEKANNTEN"-Schablonenspeicher 4K-RAM-Speicher 43 aufgezeichnet. Nun wird die Empfindlichkeit auf Spektraländerungen erhöht und neue Spektren werden immer dann aufgezeichnet, wenn eine gegen einen zweiten Schwellenwert gemessene geringfügige Änderung auftritt. Bei jeder signifikanten Änderung wird ein im Personalrechner 20 angeordneter Abtastzähler (NSAMP) aufge zählt. Diese Zählung muß ein Minimum von MINSAM, nämlich 16 unterschiedliche Spektralformen erreichen, bevor das System ein Wort als gültig erklärt, sonst wird der Schall als Hintergrundrauschen bestimmt. Jeder 5 ms-Rahmen, der keine signifikante Spektraländerung aufweist, ist ein Hinweis auf das Wortende. Vergehen 160 ms ohne Spektrumsänderung, dann wird das letzte Spektrum als wahrscheinliches Wortende erklärt und eine Musterübereinstimmungsprüfung beginnt. Ein Flußdiagramm dieses Verfahrens ist in Fig. 6 veranschaulicht.

- 1 Der Ablauf beginnt mit einem Zustand 47, der mit "Ruhezu-
stand, kein Wort" bezeichnet ist. Der Abtastzähler (NSAMP)
beginnt bei Null zu zählen und wenn die Differenz zwischen
dem augenblicklichen Spektrum und dem Hintergrundrauschen
5 den Schwellenwert t_1 überschreitet, dann läuft das Verfah-
ren zum Zustand 48, der mit "möglicher Wortbeginn" be-
zeichnet ist. Überschreitet die Differenz zwischen dem
augenblicklichen und dem letzten Spektrum nicht den zwei-
ten Schwellenwert t_2 , dann geht der Ablauf zum Kreis 49,
10 der mit "NSCNG = NSCHG + 1" bezeichnet ist. Ist die Zeit
seit der letzten Spektraländerung kurz, dann kehrt der
Ablauf zurück zum Zustand 48, um die Messung von Spektral-
änderungen zwischen dem augenblicklichen und dem letzten
Spektrum fortzusetzen. Ist die Zeit seit der letzten Spek-
15 traländerung lang - bei dem vorliegenden Ausführungsbei-
spiel etwa 160 ms - dann folgt im Ablauf der Zustand 50,
der mit "mögliches Wortende" bezeichnet ist. Ist die Zäh-
lung in dem Abtastzähler geringer als 16, dann kehrt der
Ablauf zurück zum Zustand 47 und beginnt erneut und die
20 Spektraländerungen werden als zu kurz für ein Wort be-
trachtet, so daß sie Hintergrundrauschen darstellen müs-
sen. Überschreitet die Zählung des Abtastzählers den Wert 16,
dann folgt der Zustand 52, mit "Wortende, stelle Über-
einstimmung des Musters mit Ausgangswert her". Somit stellt
25 das System fest, daß ein Wort gesprochen wurde und es be-
ginnt die Musterübereinstimmungsprüfung.

- Sobald die Spektraländerung zwischen dem augenblicklichen
und letzten Spektrum den Schwellenwert t_2 überschreitet,
30 folgt Zustand 51, der mit "Bringe signifikantes Spektral-
modell auf neuesten Stand" beschrieben ist. Ist der Ein-
gangspuffer des Abtastzählers NSAMP nicht gefüllt, dann
kehrt der Ablauf zum Zustand 48 für die nächste 5 ms-Ab-
tastung zurück. Wird der Eingangspuffer des Abtastzählers
35 NSAMP bei einer großen Spektraländerung gefüllt, dann geht
der Ablauf direkt zum Zustand 50, wo dies als Wortende
bestimmt wird und es folgt Zustand 52, in dem die Her-
stellung der Musterübereinstimmung beginnt. Wird der Ein-

1 gangspuffer des Abtastzählers NSAMP aufgrund eines kurzen
Wortes nicht gefüllt, dann ergeben sich schließlich keine
Spektraländerungen in den Abtastungen und der Ablauf geht
zum Zustand 49 wie zuvor beschrieben.

5

Bei dem Terminal des vorliegenden Ausführungsbeispiels
ist eine vorbestimmte Anzahl von Sprechern autorisiert,
das Terminal zu verwenden und Beispiele vorbestimmter
Wörter und Phrasen, wie sie von jedem Sprecher gesprochen
10 wurden, sind in dem Austauschspeicherspeicher des Rech-
ners 20 gespeichert. Die Worterkennungseinheit 16 dient
zur Unterstützung bei der Sprechererkennung bei einer et-
was vereinfachten Ausführungsform. Wenn ein spezieller
Sprecher auf das System zugreift, identifiziert er sich
15 sprachlich durch Name, Stand und Personalnummer oder ir-
gendeine andere Identifizierungszahl. Der Beginn und das
Ende jedes Wortes wird von der Worterkennungseinheit 16
festgestellt, die den Personalrechner 20 von dem gespro-
chenen Wort in Kenntnis setzt. Eine elektrische Darstel-
20 lung von LPC-Parameterdaten der LPC-Analysiereinheit 15
wird über den gesprochenen Bereich jedes Wortes gemit-
telt, dann in der CPU 18 mit einem gespeicherten Beispiel
vom Rechner 20 zur Übereinstimmung gebracht. Die Ergebnis-
se der Übereinstimmungsprüfung werden mit einem Schwellen-
25 wert verglichen, um eine Entscheidung über die Identität
des Sprechers herbeizuführen.

Während der Benutzer das System weiter verwendet, erkennt
der Rechner 20 Stellen in Sätzen, wo die Anzahl möglicher
30 nächster Wörter verhältnismäßig gering ist, wie dies jetzt
beschrieben wird. An diesen syntaktischen Knoten lädt der
Personalrechner 20 Muster oder Schablonen, d.h. gespei-
cherte Modelle von Wörtern aller Sprecher für diese näch-
sten möglichen Wörter. Beim nächsten gesprochenen Wort er-
35 kennt die Worterkennungseinheit diese Tatsache und ver-
gleicht die in das System geladenen Muster mit der Dar-
stellung des gerade gesprochenen Wortes. Die Worterkennungs-

- 1 einheit zeigt das gesprochene Wort an der Sichtanzeige des
Rechners 20 und auch den Sprecher an. Der Rechner 20 be-
sitzt einen Abstimmzähler für jeden der möglichen autori-
sierten Sprecher. Der Zähler des angezeigten Sprechers
5 wird mit jedem erkannten Wort aufgezählt bis zu einem Ma-
ximum von 25 und die Zähler aller nichtangezeigten Spre-
cher werden abwärts gezählt bis zu einer unteren Grenze
von Null. Wird beispielsweise eine Geheiminformation an-
gefordert, dann werden die Zähler geprüft und als identi-
10 fizierter Sprecher derjenige bestimmt, dessen Zählung über
15 liegt, während alle anderen Zählungen unter 8 liegen
müssen. Werden diese Bedingungen nicht erfüllt, dann wird
die Geheiminformation abgelehnt. Das System kann den Be-
nutzer im weiteren Identifikationsalgorithmus auffordern,
15 beliebige Wörter zu sprechen, bis ein eindeutiger Gewin-
ner mit entsprechendem Abstand angezeigt wird, oder das
System kann in seinem normalen Ablauf fortfahren und zu
einem späteren Zeitpunkt die Information nochmals anfor-
dern. Das System kann eine Änderung des Sprechers inner-
20 halb von maximal 10 Wörtern erkennen. Auch ist der Spre-
cheridentifikationsalgorithmus dem Benutzer im allgemei-
nen erkennbar und er weiß nicht, daß seine Stimme während
des normalen Ablaufs analysiert wird.
- 25 Die Verifikationssystemsoftware wird von den Austausch-
platten des Rechners 20 geladen und dieses Laden wird
durch Prüfsummentests verifiziert. Als nächstes werden
statistische Muster jedes bekannten Sprechers ebenfalls
geladen. Während der unbekannte Sprecher spricht, werden
30 Langzeitstatistiken der LPC-Reflexionskoeffizienten in
Echtzeit über die letzten 30 Sekunden der Sprache berech-
net. Diese Statistiken schließen eine Mittelwert- und
Standardabweichung der Tonhöhe und die ersten 10 Reflexions-
koeffizienten ein. Am Ende jedes Wortes, wie es durch die
35 Worterkennungseinheit 16 bestimmt wurde, berechnet die
CPU 18 die Mehalanobisabstandsmetrik zwischen dem unbekann-
ten Wort und dem Muster jedes Sprechers. Der Mehalanobis-

1 abstand gewichtet den Abstand mit der Fähigkeit jedes
Messungs-Eigenfektors, um den bekannten Sprecher von der
allgemeinen Bevölkerung zu unterscheiden. Schließlich in-
formiert die CPU über den Sprecher mit der besten Überein-
5 stimmung und bestimmt die Genauigkeit der Schätzung durch
den Mahalanobisabstand unter Verhältnisbildung zur Stand-
ardabweichung dieses Sprechers und durch das Verhältnis
zu der nächstbesten Übereinstimmung. Zweideutige Ergebnis-
se d.h., wenn die Übereinstimmung innerhalb eines vorbe-
10 stimmten Unsicherheitsbereichs liegt, bewirken, daß das
System eine Entscheidung zurückstellt, wodurch die Ge-
nauigkeit erhöht wird. Schließlich wird am Ende des Nach-
richtenaustausches dem Sprecher die Möglichkeit gegeben,
sein Stimmenmodell durch die zusammengesetzten Statisti-
15 ken dieses Nachrichtenaustausches auf den neuesten Stand
zu bringen.

Die LPC-Analysiereinheit 15 und die CPU 18 besitzen auch
eine Trainingsarbeitsweise bei der sich diese Statistiken
20 eines gegebenen Sprechers ergeben und in der die Eigen-
faktoren und Werte des Modells dieses Sprechers berechnet
werden. Das System kann diese Daten zur Speicherung auf
den Austauschplatten des Rechners 20 aufwärts laden.
Während die Worterkennungseinheit 16 als getrennte Ein-
25 heit des Systems veranschaulicht wird, weist der Fachmann,
daß sie in einfacher Weise auch in die LPC-Analysierein-
heit 15 oder die CPU 18 eingefügt sein kann, so daß die-
se Einheiten die Aufgaben der Erkennung des Beginns und
Endes eines Wortes, des spezifischen Wortes und des Spre-
30 chers durchführen können. Auch können Schablonen oder
Wortmodelle, die allgemein repräsentativ für jedes speziel-
le zu erkennende Wort sind, an Stelle eines Wortmodells
für jedes von jedem Sprecher gesprochene zu erkennende
Wort verwendet werden, wobei nur die speziellen Wörter
35 durch die Einrichtung erkannt würden und nicht jedoch
jeder spezielle Sprecher.

Ein typisches Beispiel einer militärischen Verwendung des

1 vorliegenden Systems sei nun in Verbindung mit den Fig. 7
und 8 erläutert. Bei dieser speziellen Ausführungsform
ist das System so aufgebaut, daß es den Verwender mit ein-
bezieht, ein geographisches Truppenmodell, Nachschub und
5 geographische Umwelt auf den neuesten Stand zu bringen.
Bei der grundsätzlichen Situation dieses Ausführungsbei-
spiels fordert der Benutzer Information von dem Terminal
an und, falls er richtig erkannt und geprüft wurde, wird
die Information von einer entfernten Quelle gegeben. Es
10 sei für dieses spezielle Ausführungsbeispiel angenommen,
daß das System um einen halben Bildschirm nach links, rechts
oben oder unten schwenken kann, oder nach Norden, Süden
Osten oder Westen bei n-Meilen. Es soll ferner die Fähig-
keit besitzen, eine Fokusierte oder eine breitere Darstel-
15 lung zu bieten, und zeigt wesentliche geographische Merk-
male, etwa eines Landesstaates einer Stadt von Gren-
zen, Straßen oder Hügel an. Bei der speziellen Anwendung
des Systems werden 55 Wörter und ein Syntaxnetzwerk mit
semantischen Zuordnungen zu jedem Knoten des Netzwerks
20 verwendet, wie dies Fig. 7 veranschaulicht. Ein Syntax-
netzwerk leitet interaktiv die Auswahl von möglichen,
nächsten Wörtern von allen dem System bekannten Wörtern
im Kontext aller Sätze, die das System versteht. Der Spre-
cher kann jederzeit sagen "Löschen" um einen neuen Satz
25 zu beginnen, oder er kann sagen "Auslöschen" um in ei-
nem Satz ein Wort zu ersetzen. Wörter wie "UH, THE", Atem-
geräusch und Zungenschlagen sind Modellwörter, die ge-
speichert werden und die von dem System absichtlich igno-
riert werden. Das System hilft dem Benutzer interaktiv,
30 wenn dieser spricht. Erwartet das System von ihm, daß er
einen Satz beginnt, d.h., wenn die Worterkennungseinheit
16 den Anfang eines ersten Wortes feststellt, dann listet
es alle möglichen ersten Wörter des Satzes auf, wie dies
in Fig. 8 A angegeben ist. Nach Sprechen des ersten Wor-
tes wird auf dem Schirm das festgestellte Wort angezeigt
35 und es werden alle möglichen zweiten Wörter gemäß Fig. 8B
aufgelistet. Dies setzt sich fort bis zum Ende des Satzes,
wenn die Daten für eine Übertragung über dem Extremschmal-

1 band Nachrichtenkanal zusammengesetzt werden. Der Sprecher
kann mit der Zeit sehen, welche nächsten Wörter erwartet
werden. Der Rechner 20 überwacht die Genauigkeit der Wort-
5 übereinstimmungen. Fällt irgendein Wort unter einen adap-
tiven Schwellenwert, dann wiederholt die Synthetisierein-
heit 22 den Satz und fragt nach fizierung vor der
Durchführung. Werden alle Wörter ganz klar erkannt, dann
gibt die Synthetisiereinheit 22 den Satz nach Vervollständi-
10 gung als Echo wieder, während der Rechner die Nachricht
aussendet."

Nach Verarbeitung jedes gesprochenen Wortes wird dieses
in den Speicher im Rechner 20 gebracht, wo die gesamte
Nachricht in ein Digitalsignal für eine minimale oder
15 fast minimale Anzahl von Bits codiert wird. Die Wörter
können in codierter Form gespeichert werden, so daß sich
der erforderliche Speicherplatz reduziert. Da das System
eine vorbestimmte Anzahl von Wörtern enthält, die es er-
kennen kann, d.h., eine vorbestimmte Anzahl von Wortmo-
20 dellen oder Mustern, so kann die Codierung in einer speziel-
len Nummer für jedes der Wörter bestehen. So kann im Bei-
spiel der Fig. 8 den Wörtern "shift focus " die Nr. 12
und dem Wort "south" die Nr. 18 zugeordnet werden, während
die Ziff. 2 durch die Nummer 21 dargestellt wird usw. Da
25 diese Wörter durch die gleichen Nummern in dem entfernten
Terminal 12 dargestellt werden, wandelt der Personalrech-
ner 20 diese Nummern in ein Digitalsignal um und überträgt
das Signal zu dem entfernten Terminal 12, wo das Signal
in Nummern und dann in Wörter zurückgewandelt wird.

30

Ein zweites Codierungsverfahren, das bei dem vorliegen-
den Ausführungsbeispiel angewandt wird, besteht darin,
jeden Buchstaben jedes Wortes in der ASC II-Codierung zu
codieren. Dieses Codierungsverfahren hat einige Vorteile,
35 obwohl es einige wenige Bits mehr pro Wort benötigt. Ei-
ner dieser Vorteile besteht darin, daß das ausgesandte
Signal direkt zu den meisten heutigen elektrisch arbeiten-
den Druckvorrichtungen übertragen werden kann. In der ASC

1 II Codierung wird jeder Buchstabe durch 8 Bits dargestellt.
Wenn somit die Musternachricht der Fig. 8 (shift focus
south 22 miles" ist, dann ist die für die Übertragung die-
5 ser Nachricht in der ASC II Codierung erforderliche Bit-
zahl gleich 260. Werden 20 Bits zur Beschreibung von Ei-
genschaften der Stimme des Sprechers verwendet und er-
fordern Synchronisationsfehlererkennung und Steuersignale
weitere 30 Bits, dann ist die vollständige Nachricht etwa
10 310 Bits lang. Es ist somit möglich eine Nachricht mit ei-
ner Länge von etwa 4 Sekunden und mit 310 Bits, d.h., mit
etwa 77 Bits pro Sekunde zu übertragen.

Wird wie zuvor beschrieben ein Codierungssystem verwendet,
bei dem jedem Wort eine spezielle Nummer zugeteilt ist,
15 dann ist die Situation folgende: nimmt man an, daß die
gesprochene Nachricht eine von 100 möglichen Nachrichten-
typen mit jeweils gleicher Wahrscheinlichkeit ist, dann
sind 7 Bits erforderlich um, um den grammatikalischen
Aufbau der Nachricht zu beschreiben. Werden 20 auswähl-
20 bare Wörter in dem System gespeichert die ausgewählt wer-
den können, um verschiedene Positionen in der Nachricht
einzunehmen, dann definieren 8 Bits welches Wort in je-
der gewünschten Position in der Nachricht verwendet wur-
de. Für die Musternachricht, wie Sie zuvor angegeben wur-
25 de, nämlich für "shift focus south 22 miles" definieren
7 Bits die Nachricht Syntax, 40 Bits definieren die 5
auswählbaren Wörtern an Positionen innerhalb der Nachricht,
wo eines von mehreren Wörtern ausgewählt werden kann,
und etwa 20 Bits können die Eigenschaften der Stimme der
30 Sprecher angeben, so daß sich eine Gesamtzahl von 67 Bits
ergibt. Werden wiederum etwa 30 Bits für die Synchronisa-
tionsfehlerkorrektur und Steuersignale angesetzt, dann
umfaßt die gesamte Nachricht etwa 97 Bits oder etwa 25
Bits pro Sekunde.

35 Die Synthesisiereinheit 22 des vorliegenden Ausführungs-
beispiels ist im Handel erhältlich und wird von der Firma

1 Mikromint Inc. als Mikrovoxsynthesizer vertrieben. Der
 Fachmann erkennt selbstverständlich, daß die LPC-Analy-
 siereinheit 15 einen Synthetisierer aufweist, (vgl. Fig. 2)
 und an Stelle der Synthetisiereinheit 22 verwendet, wenn
 5 die Sprechererkennung in dem System eingeschlossen ist
 und wenn es erwünscht ist, daß die syntetisierte Stimme
 der Stimme des ursprünglichen Sprechers gleicht. Die
 Synthetisiereinheit 22 wurde jedoch hier beschrieben und
 zwar der Einfachheit und des besseren Verständnis halber.
 10 Von der Beschreibung der Synthetisiereinheit 22 ergibt sich
 für den Fachmann ein vollkommenes Verständnis der Arbeits-
 weise des in der LPC-Analysiereinheit 15 vorhandenen Syn-
 thetisierers. Eine vollständigere Beschreibung des Syntheti-
 sierers der in der LPC-Analysiereinheit 15 enthalten ist,
 15 ergibt sich aus der zuvor genannten Patentanmeldung und
 aus der US-Patentanmeldung mit der Bezeichnung "Speech
 Synthesizer With Smooth Linear Interpolation", mit dem
 Aktenzeichen 267 203, eingereicht am 26 Mai 1981.

20 Die Synthetisiereinheit 22 ist ein freistehender inteligen-
 ter Mikroprozessor, der ASCII Text in gesprochenes Englisch
 umwandelt. Sie besteht aus einem M 65 02 Mikroprozessor 55,
 einer 9600 BPS UART-Einheit 57 als serielle Schnittstelle
 einem RAM-Speicher 59 mit einer Speicherkapazität von 2K
 25 Bits einem löschbaren, programmierbaren Nur-Lesespeicher
 EPROM 61 mit 8 K-Bits, einem SC01 Votrax-Stimmsyntetisier-
 er 63, einem taktenden und programmierbaren Teiler 65 und
 verschiedenen Puffern, Steuerungen und Verstärkern. Die
 Synthetisiereinheit 22 verwendet einen Algorithmus, der
 30 grammatikalisch Serieneingangsdaten in Wörter umsetzt,
 dann die englischen Ausspracheregeln verwendet und einen
 Lautstrom aus dem ausgesprochenen zu erzeugen. Dieser
 Lautstrom steuert dann den Sprachsynthetisierer 63. Der
 Sprachsynthetisierer 63 besitzt einen ROM-Speicher der
 35 Laute als eine Folge von 1 bis 4 Tönen in ständigem Zu-
 stand von spezifischer Dauer und mit spezifischem Spek-
 trum erzeugt. Die Funktion der Synthetisiereinheit 22 be-
 ruht auf den Buchstaben zu Laut-Umsetzungsregeln, die

- 1 in dem Mikroprozessor 55 angewandt werden, sowie auf der Laut-Sprachensyntese in dem Sprachsynthetisierer 63. Der Mikroprozessor 55 liest bis zu 1500 Zeichen in seinen internen Seitenpuffer von der seriellen Schnittstelle 57.
- 5 Er identifiziert Phrasengruppen durch ihre Punctuation und Wörter durch ihre Zwischenraumbegrenzer. Er verwendet die Phrasengruppengrenzen um eine geeignete deklarative oder fragende Tonhöhen- und Dauerbeugung auf die Phrase anzuwenden. Pro Wort wird jedes Zeichen von links nach
- 10 rechts über das Wort abgetastet. Wird ein Zeichen gefunden, bei dem die linken und rechten Kontexterfordernisse, (benachbarte Zeichen) erfüllt sind, dann wird die erste anwendbare Regel für das Zeichen verwendet, um es in einen Laut umzusetzen.
- 15 Der Sprachsynthetisierer 63 ist ein CMOS-Typ, der aus einem digitalen Codeumsetzer und einem elektronischen Modell der Vokalspur besteht. Intern ist eine Lautsteuerung vorgesehen, die eine 6-Bit-Laut- und 2-Bit-Tonhöhencodierung in eine
- 20 Matrix von spektralen Parametern umsetzt, die das Vokalspurmodell zur Synthetisierung der Sprache einstellt. Die Ausgangstonhöhe der Laute wird durch die Frequenz des getakteten Teiler 65 abgegebenen Taktsignal gesteuert. Feine Schwankungen der Tonhöhe können induziert werden, um
- 25 eine Beugung hinzuzufügen, was verhindert, daß die synthetisierte Stimme monoton und maschinell klingt. Während der vorliegende Algorhythmus einen englischen Text in Sprache umwandelt, ist es für den Fachmann verständlich, daß die Sprachalgorhythmus genauso in anderen Sprachen geschrieben sein können. 64 Laute definieren die englische Sprache
- 30 und jeder Laut wird durch eine 6-Bit-Codierung gekennzeichnet, die von dem Mikroprozessor 55 an den Sprachsynthetisierer 63 angelegt wird. Die Lautsteuerung setzt dann die Bits in die zuvor erwähnten Spektralparameter um.
- 35 Damit die synthetisierte Sprache möglichst gut dem identifizierten ursprünglichen Sprecher gleicht, können verschiedene Codierungen senderseitig zu dem empfangenden Gerät

1 übertragen werden, wobei Daten über die spezielle Aus-
 sprache des Sprechers bezüglich dieser Worte beinhalten.
 Dies kann sehr einfach dadurch erreicht werden, daß eine
 Sprecheridentifikationscodierung ausgesandt wird, die der
 5 Empfänger zum Aufsuchen der Vokalspurlänge und des mittler-
 en Tonhöhenbereichs verwendet. Alternativ dazu kann der
 Sender auch Polynomkoeffizienten aussenden, die die Tonhöhen-
 kontur über der Länge des Satzes beschreibt, sowie einen
 Vokalspurlängenmodifizierer. Diese Polynomkoeffizienten
 10 ermöglichen, daß der richtige Tonhöhenbereich, Tonhöhen-
 abfall und die Betonung mit sehr wenigen Bits übertragen
 werden. Der Vokalspurlängenmodifizierer ermöglicht es
 dem Syntetisierer eine Polynominterpolation der LPC-Re-
 flektionskoeffizient durchzuführen, wodurch die
 15 Vokalspur länger oder kürzer gemacht werden kann als bei
 dem gespeicherten Muster, das bei den Buchstaben- Zu- Ton-
 Regeln verwendet wird.

Es wurde somit ein Extremschmalband-Übertragungssystem
 20 offenbart, bei dem jedes Terminal menschliche Stimme in
 Digitalsignale mit einer Geschwindigkeit von weniger als
 300 Bits pro Sekunde umsetzt. Das Terminal besitzt fer-
 ner die Fähigkeit Digitalsignale zu empfangen, die re-
 präsentativ für eine menschliche Stimme sind, und die
 25 menschliche Stimme mit den gleichen Eigenschaften wie
 die des ursprünglichen Sprechers zu synthetisieren. Außer-
 dem besitzt jedes Terminal die Fähigkeit Wörter und den
 speziellen Sprecher mit sehr hoher Genauigkeit zu erken-
 nen.

30

35

-23-

- Leerseite -

This Page Blank (uspio,

NACHGEREICHT

Num.
Int. Cl.
Anmeldetag:
Offenlegungstag:

34 16 238
G 10 L 1/00
2. Mai 1984
20. Dezember 1984

-31-

1/8

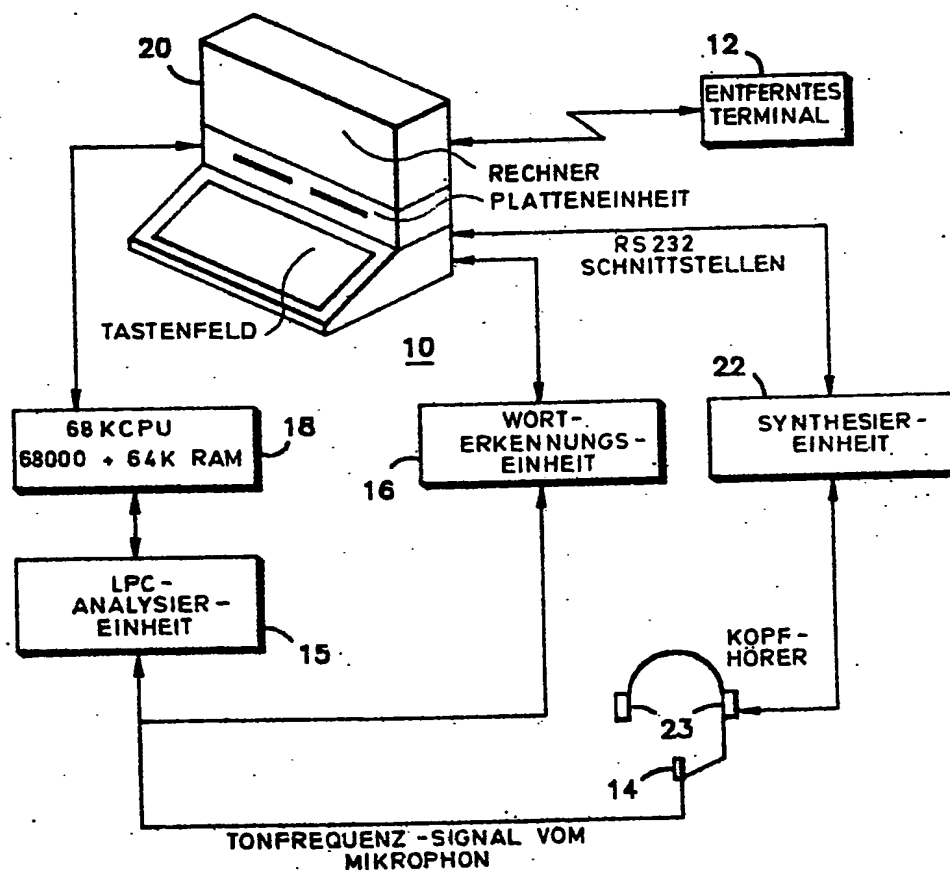


FIG. 1

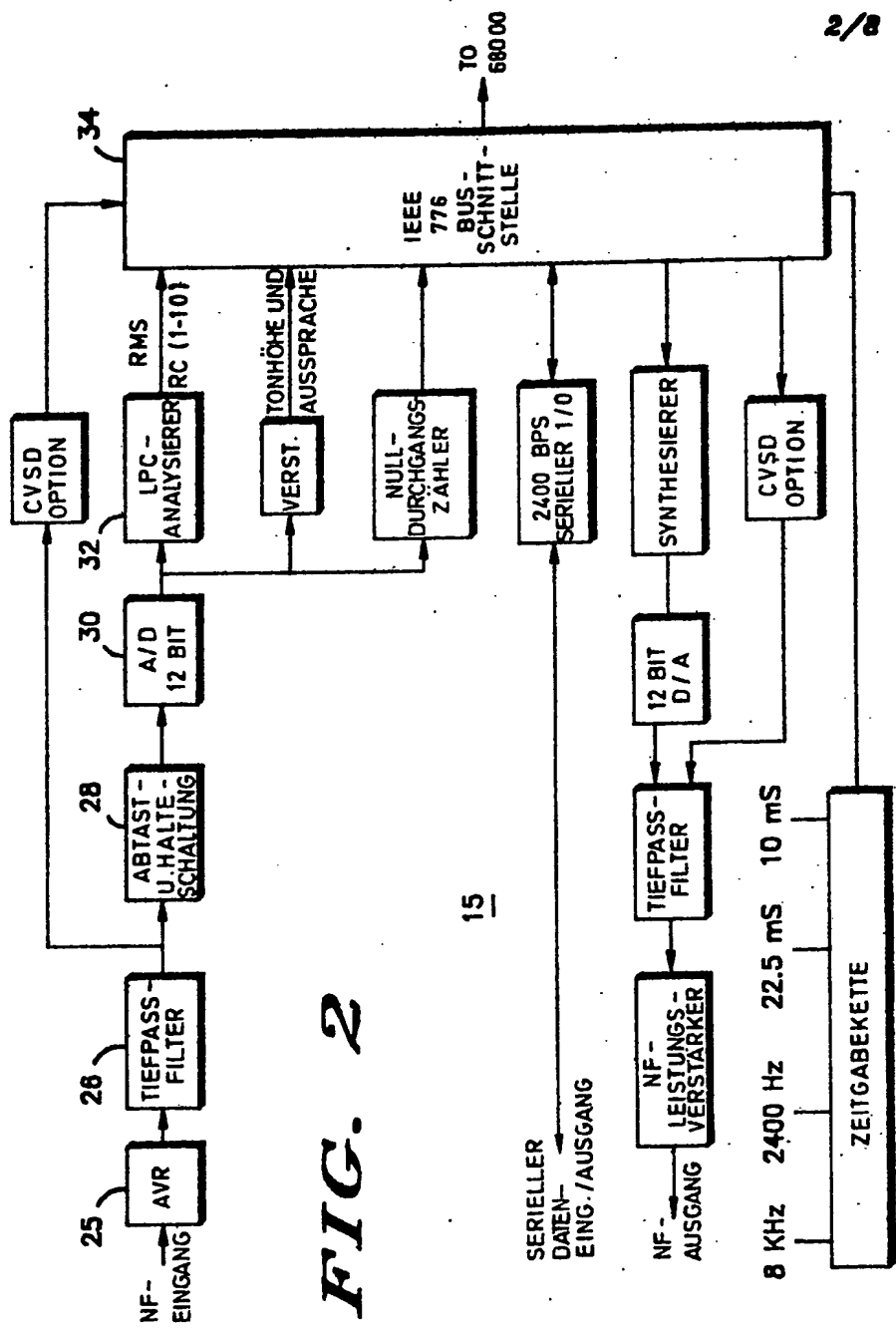
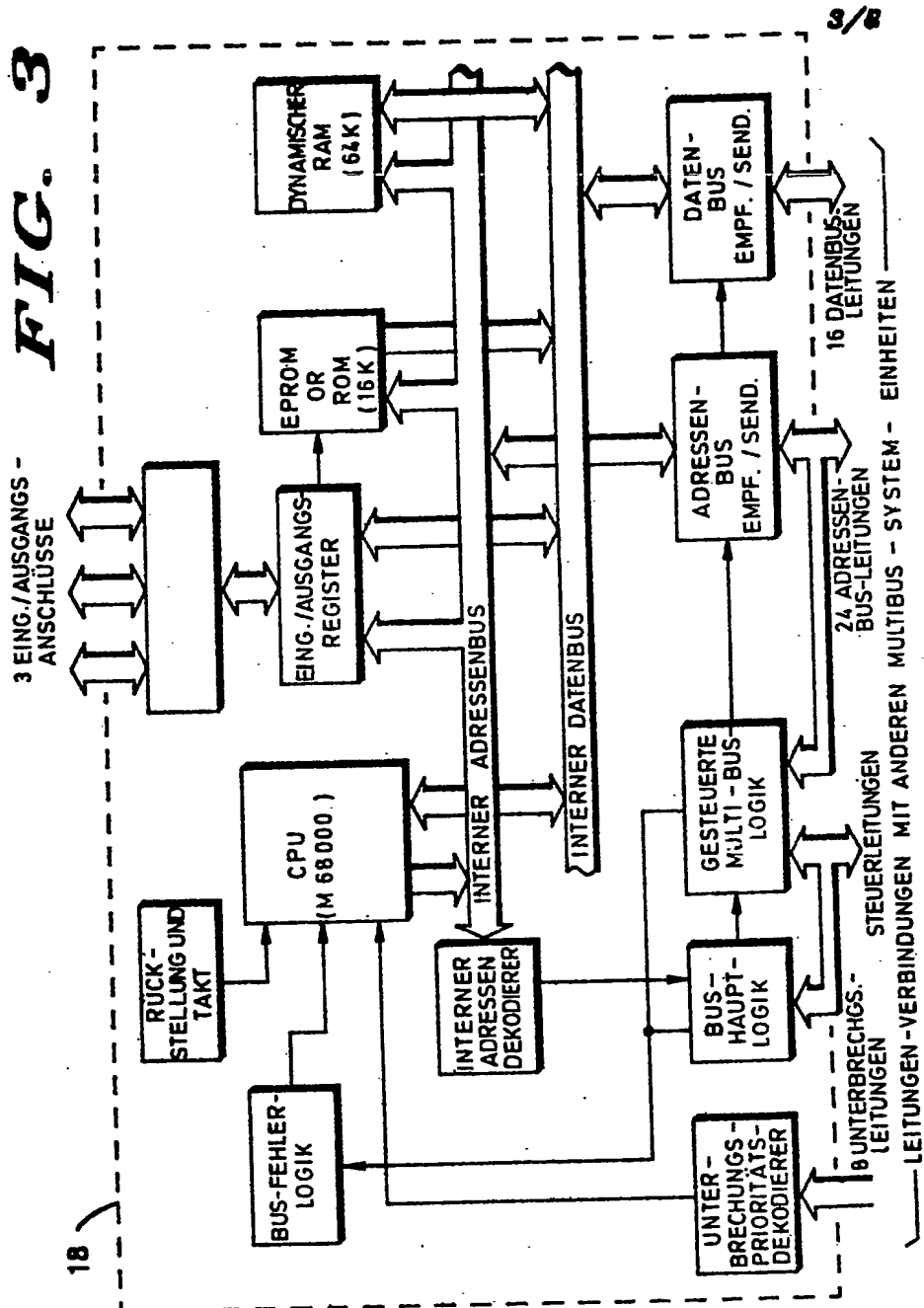


FIG. 2

FIG. 3



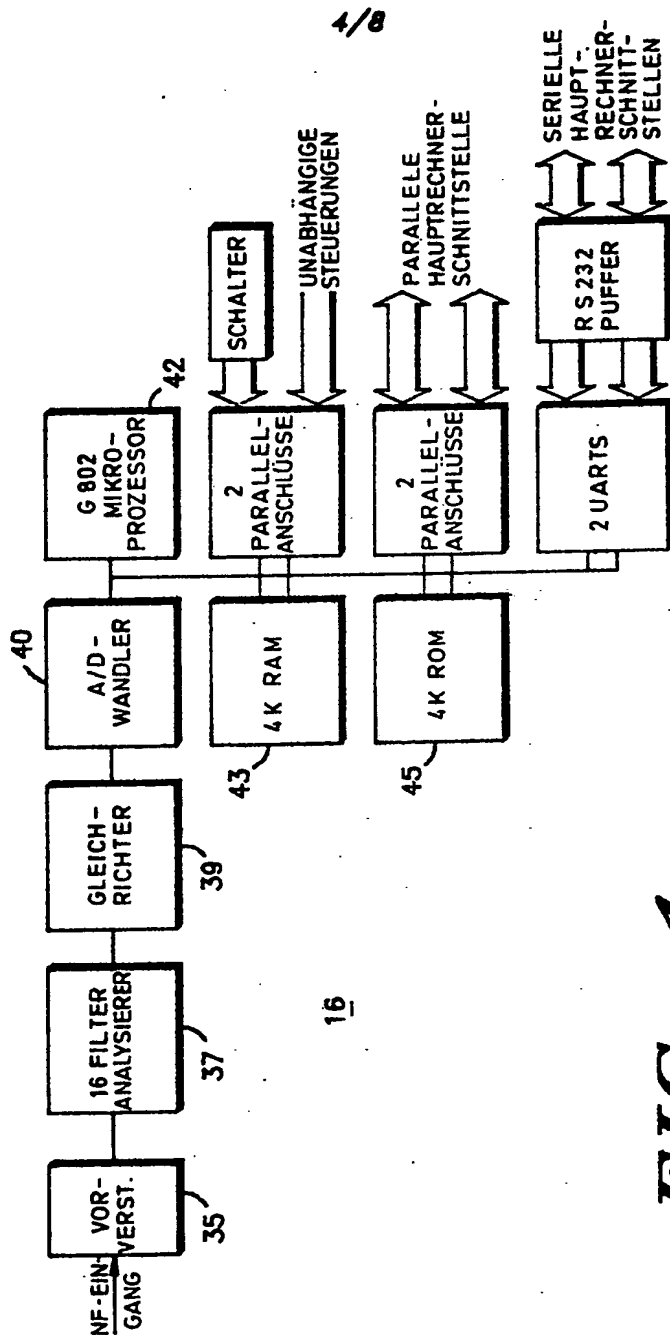


FIG. 4

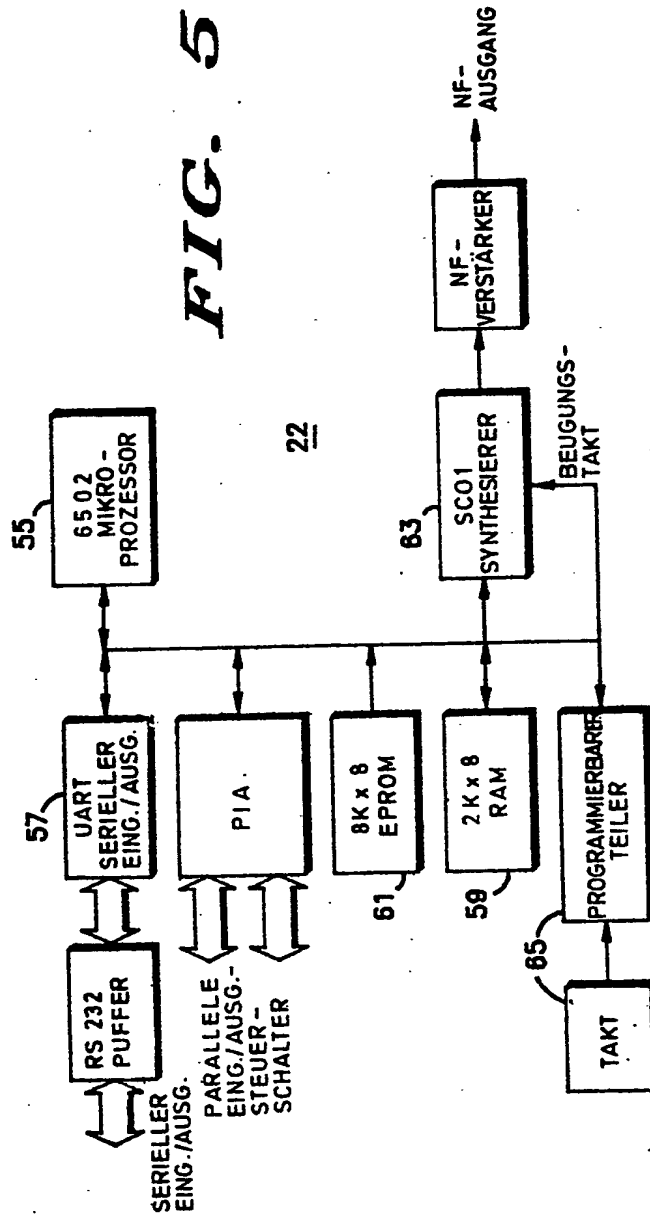
NACHSCHREIBT

27

3416238

5/8

FIG. 5



NACHGESCHREIBT

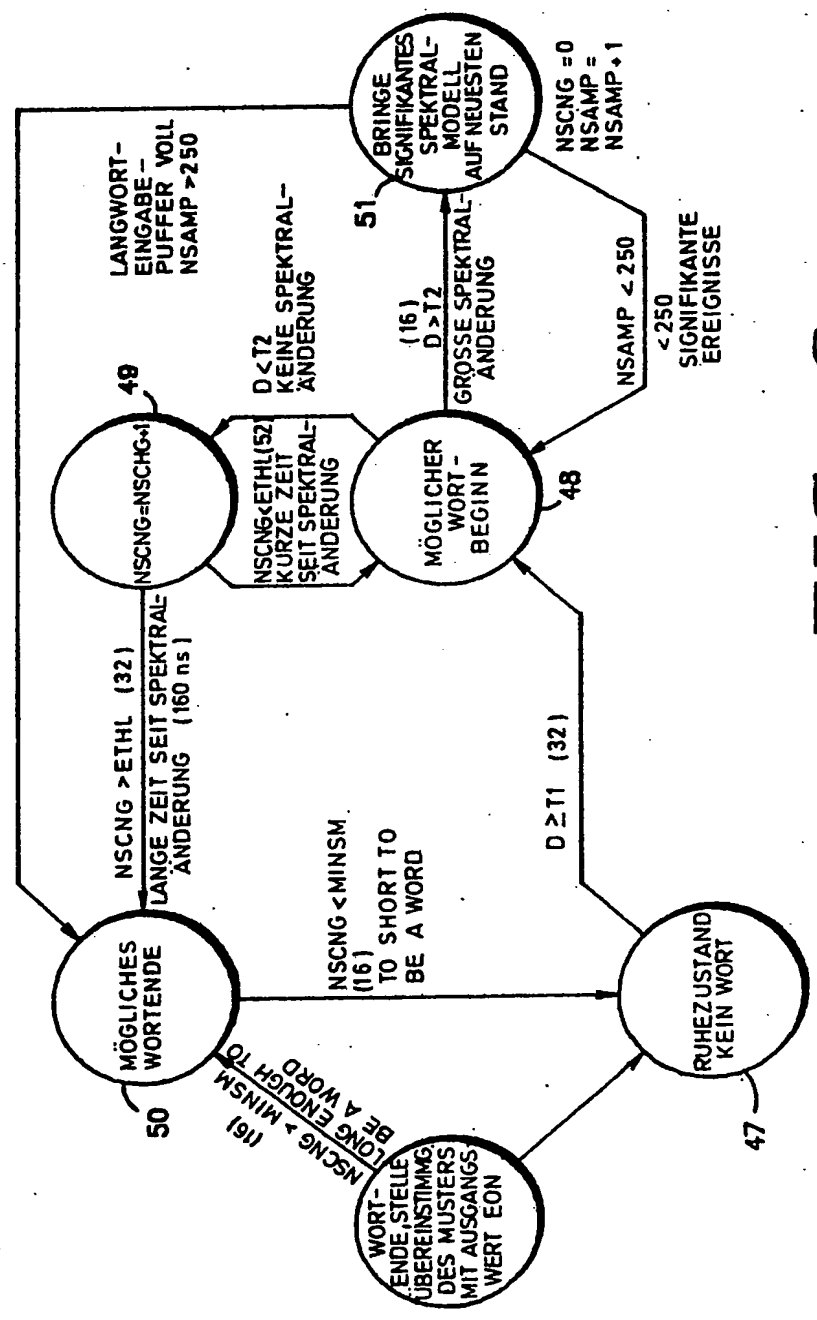
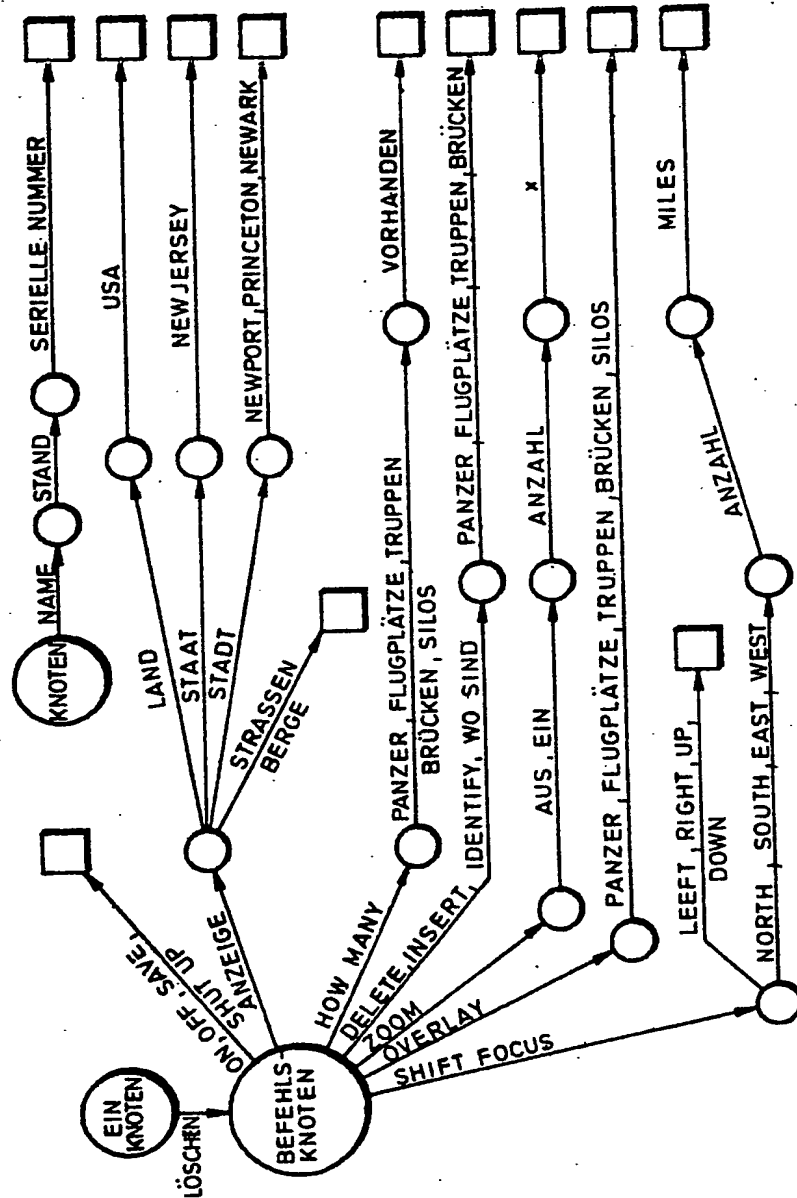


FIG. 6

7/8

FIG. 2

BEMERKUNGEN: 1. LÖSCHEN KANN BEI JEDEM KNOTEN VERWENDET
WERDEN, AUSSER BEI LETZTEM KNOTEN- UM
EINEN KNOTEN IM ENTSCHEIDUNGSBAUM ZU ER-
SETZEN
2. "UH" UND "THE" WERDEN IMMER IGNORIERT



NACHGEREICHT

8/8

ON
OFF
SAVE
SHUTUP
DISPLAY
HOW MANY
DELETE
INSERT
IDENTIFY
ZOOM
OVERLAY
SHIFTFOCUS

SHIFT FOCUS LEFT
RIGHT
UP
DOWN
NORTH
SOUTH
EAST
WEST

SHIFT FOCUS SOUTH 0
1
2
3
4
5
6
7
8
9
POINT

SHIFT FOCUS SOUTH 2 0
1
2
3
4
5
6
7
8
9
POINT
MILES

FIG.
8